

CLAIMS

1. A method for processing a data stream comprising:
 - receiving a data segment;
 - determining whether the data segment has been previously stored; and
 - 5 in the event that the data segment is determined not to have been
previously stored, generating a unique identifier for specifying the data segment
in a representation of the data stream.
2. A method for processing a data stream as recited in Claim 1 wherein determining
whether the data segment has been previously stored includes generating a content
10 derived summary.
3. A method for processing a data stream as recited in Claim 1 wherein determining
whether the data segment has been previously stored includes generating a content
derived summary for the data segment; and the content derived summary is a fingerprint.
4. A method for processing a data stream as recited in Claim 1 wherein determining
15 whether the data segment has been previously stored includes looking up a content
derived summary for the data segment; and the content derived summary is the data
segment.
5. A method for processing a data stream as recited in Claim 1 wherein determining
whether the data segment has been previously stored includes generating a content
20 derived summary for the data segment; and locating the content derived summary in a
content derived summary storage.

6. A method for processing a data stream as recited in Claim 1 wherein determining whether the data segment has been previously stored includes locating the data segment in a data segment storage.

7. A method for processing a data stream as recited in Claim 1 wherein in the event
5 that the data segment is determined not to have been previously stored, further including storing the data segment in a data segment storage location.

8. A method for processing a data stream as recited in Claim 1 wherein:
determining whether the data segment has been previously stored includes
generating a content derived summary for the data segment;

10 in the event that the data segment is determined not to have been
previously stored, further including:

storing the data segment in a data segment storage location; and
updating a data structure for storing the content derived summary,
the unique identifier, and the data segment storage location.

15 9. A method for processing a data stream as recited in Claim 1 wherein:
determining whether the data segment has been previously stored includes
generating a content derived summary for the data segment;

in the event that the data segment is determined not to have been
previously stored, further including:

20 storing the data segment in a data segment storage location; and
updating a data structure for storing the content derived summary,
the unique identifier, and the data segment storage location; wherein

the data segment storage location is accessed given the unique identifier or given the content derived summary in the data structure.

10. A method for processing a data stream as recited in Claim 1 wherein:

determining whether the data segment has been previously stored includes

5 generating a content derived summary for the data segment;

in the event that the data segment is determined not to have been previously stored, further including:

storing the data segment in a data segment storage location; and

updating a data structure for storing the content derived summary,

10 the unique identifier, and the data segment storage location; wherein

the data segment storage location is accessed given the unique identifier or given the content derived summary, using a single access of a storage device.

11. A method for processing a data stream as recited in Claim 1 wherein:

15 determining whether the data segment has been previously stored includes generating a content derived summary for the data segment;

in the event that the data segment is determined not to have been previously stored, further including:

storing the data segment in a data segment storage location; and

20 updating a data structure for storing the content derived summary,

the unique identifier, and the data segment storage location; wherein

a region of the data structure that includes the data segment storage location is accessed given the unique identifier or given the content derived summary, using a single access of a storage device.

12. A method for processing a data stream as recited in Claim 1, wherein the unique
5 identifier is a short identifier that does not depend on probability for its uniqueness.
13. A method for processing a data stream as recited in Claim 1, wherein the unique identifier is a serial number.
14. A method for processing a data stream as recited in Claim 1, wherein the unique identifier is derived from a hash value.
- 10 15. A method for processing a data stream as recited in Claim 1, wherein the unique identifier is an address of the data segment.
16. A method for processing a data stream as recited in Claim 1, wherein the unique identifier is a shortest identifier for uniquely identifying the data segment.
17. A method for processing a data stream as recited in Claim 1, wherein determining
15 whether the data segment has been previously stored includes generating a content derived summary for the data segment; and the unique identifier is derived from the content derived summary.
18. A method for processing a data stream as recited in Claim 1, wherein determining
20 whether the data segment has been previously stored includes generating a content derived summary for the data segment; and the unique identifier includes a value derived from the content derived summary and a serial number.
19. A method for processing a data stream as recited in Claim 1, wherein the representation of the data stream is a compressed representation.

20. A method for processing a data stream as recited in Claim 1, wherein the representation of the data stream is used for reconstructing the data stream.
21. A method for processing a data stream as recited in Claim 1, wherein determining whether the data segment has been previously stored includes generating a candidate
5 identifier; and determining whether the candidate identifier has been stored previously.
22. A method for processing a data stream as recited in Claim 1, wherein:
determining whether the data segment has been previously stored includes
generating a candidate identifier; and determining whether the candidate identifier
has been stored previously;
10 generating a unique identifier for specifying the data segment includes
modifying the candidate identifier.
23. A method for processing a data stream as recited in Claim 1, wherein modifying the candidate identifier includes adding a value to the candidate identifier.
24. A method for processing a data stream as recited in Claim 1, wherein modifying
15 the candidate identifier includes combining an additional bit with the candidate identifier.
25. A method for processing a data stream as recited in Claim 1, wherein modifying the candidate identifier includes combining a plurality of bits with the candidate identifier.
26. A method for processing a data stream as recited in Claim 1, wherein the unique
20 identifier is stored in a reconstruction list.
27. A method for processing a data stream as recited in Claim 1, in the event that the data segment is determined to have been previously stored, further including locating a unique identifier previously assigned to the data segment.

28. A method for processing a data stream as recited in Claim 1, in the event that the data segment is determined to have been previously stored, further including locating a unique identifier previously assigned to the data segment; and the unique identifier is stored in a reconstruction list.

5 29. A method for processing a data stream as recited in Claim 1, further comprising:
determining whether the data segment has been previously stored; and
in the event that the data segment is determined not to have been
previously stored, storing the data segment.

30. A system for processing a data stream comprising:
10 an interface configured to receive a data segment;
a processor coupled to the interface, configured to:
determine whether the data segment has been previously stored;
and
in the event that the data segment is determined not to have been
15 previously stored, generate a unique identifier for specifying the data
segment in a representation of the data stream.

31. A computer program product for processing a data stream, the computer program product being embodied in a computer readable medium and comprising computer instructions for:
20 receiving a data segment;
determining whether the data segment has been previously stored; and

in the event that the data segment is determined not to have been previously stored, generating a unique identifier for specifying the data segment in a representation of the data stream.